Capstone Project - The Battle of the Neighborhoods

Applied Data Science Capstone by IBM/Coursera

Table of contents

Business Problem

This project aims to find the most suitable and secure location for opening some kinds of business places like **Grocery Store**, **Gas Station**,... in **Toronto**, Canada. We will use data science tools to analyse data and focus on the safest borough and explore its neighborhoods and the 10 most common venues in each neighborhood so that the best neighborhood where grocery stores are not amongst the most common venues can be selected.

Data

Based on definition of our problem, we must:

- Find the safest borough based on crime statistics
- Find the most common venues

Following data will be needed to extract the required information:

Part 1: Using a real world data set from Kaggle containing the Toronto Police Data from 2014 to 2019: A dataset consisting of the crime statistics of Toronto along with coordinates of neighborhoods, type of crime, and record time.

Part 2: Use the data from Wikipedia about neighborhoods and borough in Toronto.: Borough information will be used to map the existing data where each neighbourhood can be assigned with the right borough.

Part 3: Find the safest borough: Analyse the data and find the safest places.

Part 1: Using a real world data set from Kaggle containing the Toronto from 2014 to 2019

	X	Y	Index_	event_unique_id	occurrencedate	reporteddate	premisetype	ucr_code	ucr_ext	offence	reportedyear	reportedn
0	-79.405228	43.656982	7801	GO-20152165447	2015-12-18T03:58:00.000Z	2015-12-18T03:59:00.000Z	Commercial	1430	100	Assault	2015	December
1	-79.307907	43.778732	7802	GO-20151417245	2015-08-15T21:45:00.000Z	2015-08-17T22:11:00.000Z	Commercial	1430	100	Assault	2015	August
2	-79.225029	43.765942	7803	GO-20151421107	2015-08-16T16:00:00.000Z	2015-08-18T14:40:00.000Z	Apartment	2120	200	B&E	2015	August
3	-79.140823	43.778648	7804	GO-20152167714	2015-11-26T13:00:00.000Z	2015-12-18T13:38:00.000Z	Other	2120	200	B&E	2015	December
4	-79.288361	43.691235	7805	GO-20152169954	2015-12-18T19:50:00.000Z	2015-12-18T19:55:00.000Z	Commercial	1430	100	Assault	2015	December
...	...	...	...	...	...	...	...	...	...	...	...	...
8154	-79.331421	43.662998	15940	GO-20151776717	2015-10-15T12:45:00.000Z	2015-10-15T14:17:00.000Z	House	1430	100	Assault	2015	October
8155	-79.542381	43.752094	15941	GO-20151780161	2015-10-15T23:15:00.000Z	2015-10-15T23:29:00.000Z	Apartment	1430	100	Assault	2015	October
8156	-79.542381	43.752094	15942	GO-20151780161	2015-10-15T23:15:00.000Z	2015-10-15T23:29:00.000Z	Apartment	1430	100	Assault	2015	October
8157	-79.373062	43.663193	15943	GO-20151813077	2015-10-21T15:59:00.000Z	2015-10-21T15:59:00.000Z	Outside	1430	100	Assault	2015	October
8158	-79.296867	43.690365	15944	GO-20151814568	2015-10-21T20:00:00.000Z	2015-10-21T20:32:00.000Z	Apartment	1430	100	Assault	2015	October

After clean the data:

	Lat	Long	offence	reportedyear	Neighbourhood
0	43.656982	-79.405228	Assault	2015	University
1	43.778732	-79.307907	Assault	2015	Tam O'Shanter-Sullivan
2	43.765942	-79.225029	B&E	2015	Woburn
3	43.778648	-79.140823	B&E	2015	Centennial Scarborough
4	43.691235	-79.288361	Assault	2015	Taylor-Massey
...	...	...	...	...	...
8153	43.651897	-79.381714	Assault Bodily Harm	2015	Bay Street Corridor
8154	43.662998	-79.331421	Assault	2015	South Riverdale
8155	43.752094	-79.542381	Assault	2015	Humbermede
8156	43.752094	-79.542381	Assault	2015	Humbermede
8157	43.663193	-79.373062	Assault	2015	Cabbagetown-South St.James Town

Total Crimes in different Neighborhoods

```
Waterfront Communities-The Island     317
Bay Street Corridor                   265
Church-Yonge Corridor                 246
West Humber-Clairville                183
Moss Park                             178
                                      ...
Runnymede-Bloor West Village           15
Bayview Woods-Steeles                  15
Forest Hill South                      14
Humber Heights-Westmount               14
Maple Leaf                             13
```

Part 2: Use the data from Wikipedia about neighborhoods and boroughs in Toronto.

|	Lat	Long	Type	Year	Neighbourhood	Borough
0	43.656982	-79.405228	Assault	2015	University	Old City of Toronto
1	43.661518	-79.409767	B&E	2015	University	Old City of Toronto
2	43.666660	-79.403870	Assault	2015	University	Old City of Toronto
3	43.665916	-79.407471	Assault Bodily Harm	2015	University	Old City of Toronto
4	43.657951	-79.400047	Assault	2015	University	Old City of Toronto
...	...	...	...	...	...	...
7715	43.709774	-79.481880	Assault Bodily Harm	2015	Maple Leaf	North York
7716	43.706375	-79.487671	Theft Over	2015	Maple Leaf	North York
7717	43.707474	-79.488113	B&E W'Intent	2014	Maple Leaf	North York
7718	43.708660	-79.478065	Assault	2014	Maple Leaf	North York
7719	43.721577	-79.480972	Unlawfully In Dwelling-House	2015	Maple Leaf	North York

# Methodology

Categorized the methodologysection into two parts:

- **Exploratory Data Analysis**: Visualise the crime reports in different boroughs to idenity the safest borough and normalise the neighborhoods of that borough.
- **Modelling**: To help stakeholders choose the right neighborhood within a borough we will be clustering similar neighborhoods using K - means clustering to address this problem so as to group data based on existing venues which will help in the decision making process.
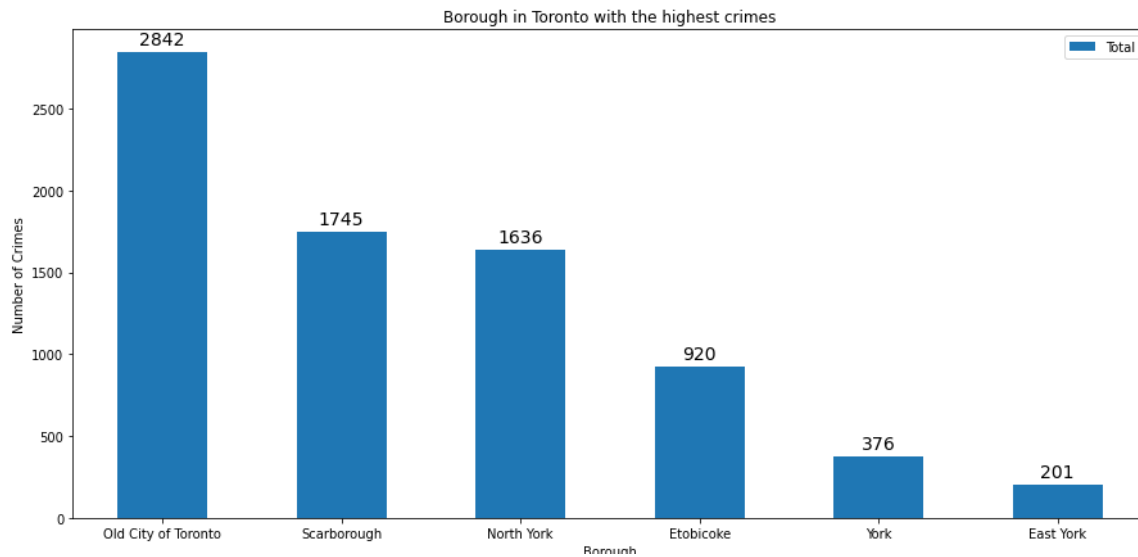
Visualize the data

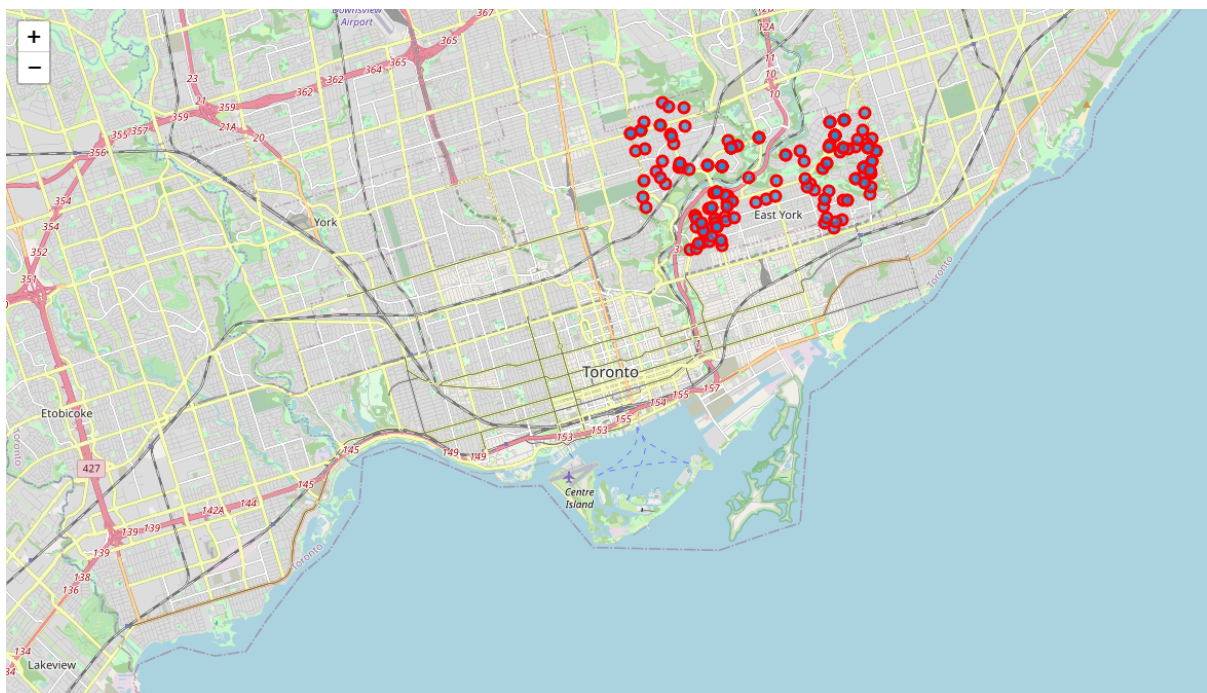*5 Neighborhoods with highest crime*
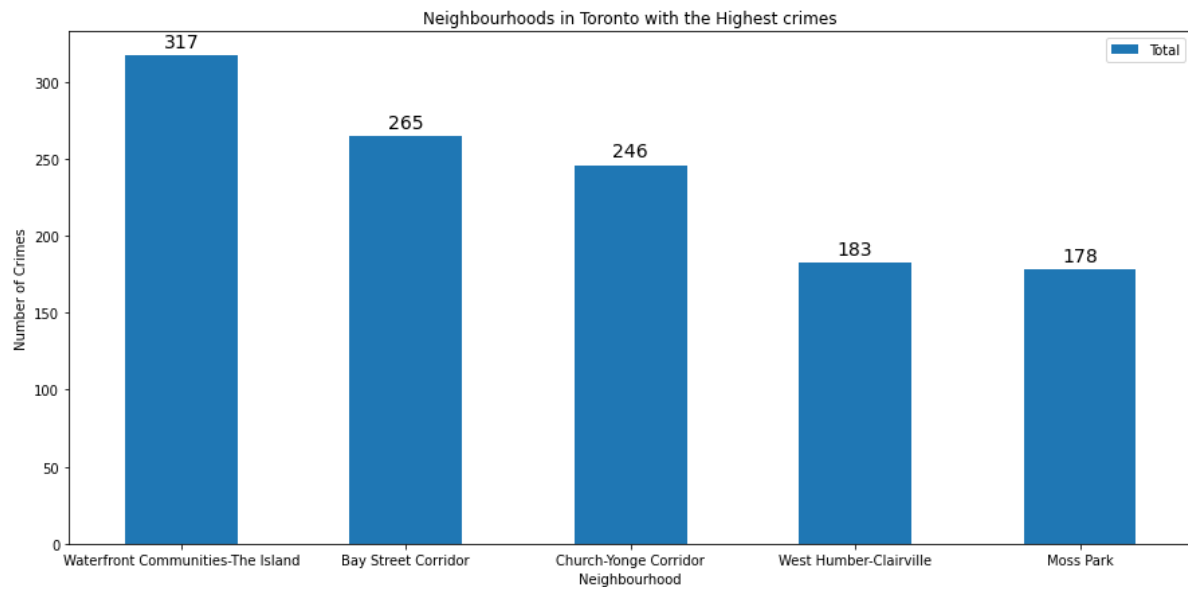
5 Neighborhoods with lowest crime
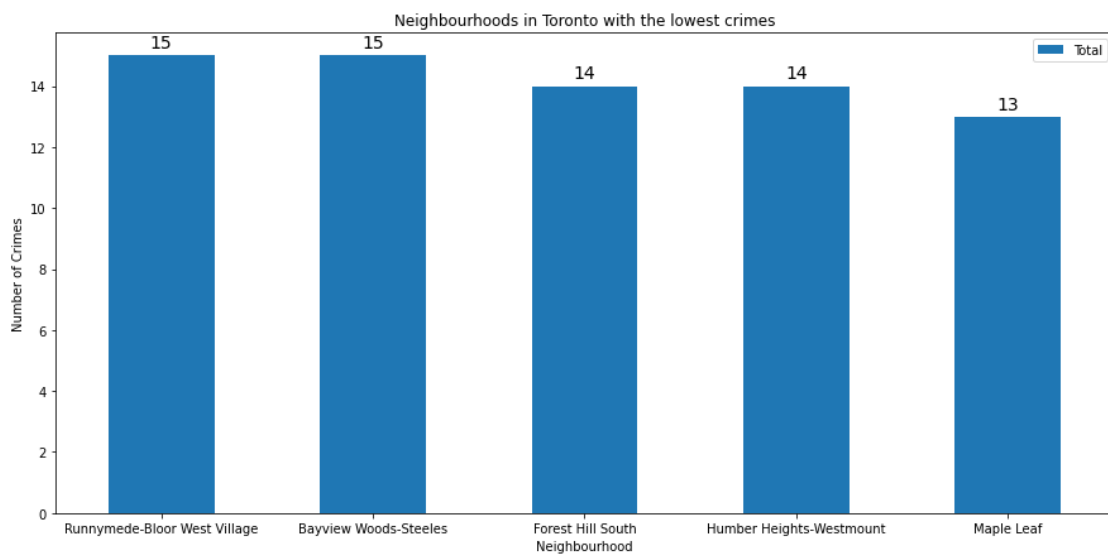
Borough with highest crime
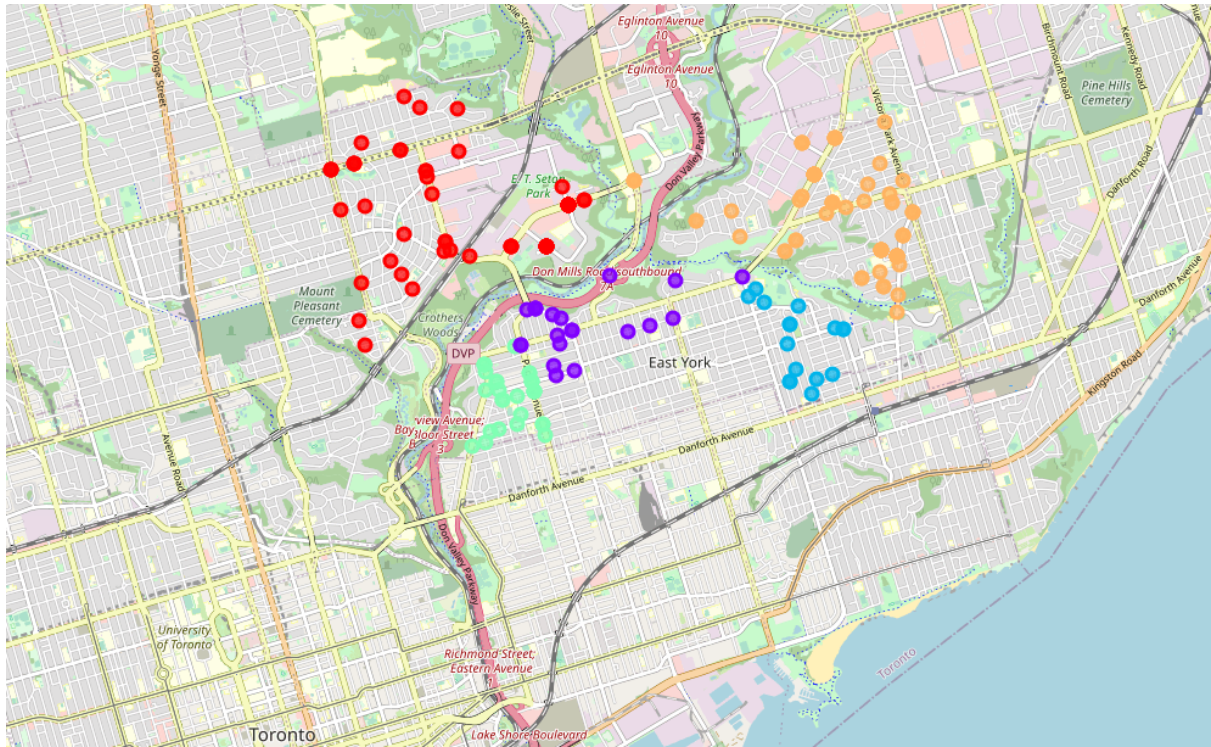
Borough in Toronto with the highest crimes

Based on exploratory data analysis it is clear that East York has the lowest crimes

Create a map to visualize neighborhoods in East York

Use K-mean clustering to cluster neighborhoods in East York

East York's map after clustering

Conclusion

We have explored the crime data to understand different types of crimes in all neighborhoods of Toronto and later categorized them into different boroughs, this helped us group the neighborhoods into boroughs and choose the safest borough first.